Data management for National Park Service– National Geographic Society BioBlitzes

Evolving biodiversity documentation

By Peter Budde and Simon Kingston

THE NATIONAL PARK SERVICE (NPS) and the National Geographic Society (NGS) have cohosted eight annual national park bioblitzes since their inception in 2007. Over time the methods and tools used to manage data from these events have evolved. Spreadsheets may have given way to smartphones and other mobile devices, but one constant has been the use of NPSpecies (https://irma.nps .gov/App/Species/Welcome), the centralized data application that documents the occurrence and status of species in national parks. NPSpecies provides a baseline of the species known to occur in a park and reflects the new knowledge gained from bioblitzes and other forms of scientific inquiry.

Information stored and managed in NPSpecies satisfies a fundamental purpose of the National Park Service to protect and maintain biological diversity in parks. Park managers, interpreters, planners, and scientists need basic information about species occurring in parks as a basis for making decisions and for working with the public, other agencies, and the scientific community.

Data management for NPS-NGS BioBlitzes occurs in three distinct stages: the pre-bioblitz buildup, the rush of activity during the bioblitz itself, and the post-bioblitz follow-up (fig. 1). Each stage has its own set of activities, but common themes are links to NPSpecies and quality assurance (QA) and quality control (QC) steps that are taken to ensure the most

> ### Bioblitz
>
> A bioblitz is commonly a 24-hour event in which teams of volunteer scientists, families, students, teachers, and other community members work together to find and identify as many species of plants, animals, microbes, fungi, and other organisms as possible. The National Geographic Society is helping conduct a bioblitz in a different national park each year during the decade leading up to the U.S. National Park Service centennial in 2016.[1]
>
> ──────────
> [1] http://www.nationalgeographic.com/explorers/projects/bioblitz/.

reliable data possible are recorded and made available.

The buildup (pre-bioblitz)

Preparation for managing data at an NPS-NGS BioBlitz involves determining the local contact (typically the Inventory and Monitoring Network data manager), designing field data sheets (fig. 2, page 22), eliciting the types of questions park staff would like to be able to answer as a result of the event, designing and developing the data entry and reporting application, and writing procedures and workflows to be used during the bioblitz.

NPSpecies
The goal of NPSpecies is to be the authoritative source of species that occur in a park; however, many park species lists were last updated and reviewed during a certification process that occurred several years ago. Before the bioblitz, in order to make sure that a park's species list reflects the current state of knowledge for organisms that occur in a park, species records from inventories, incidental observations,
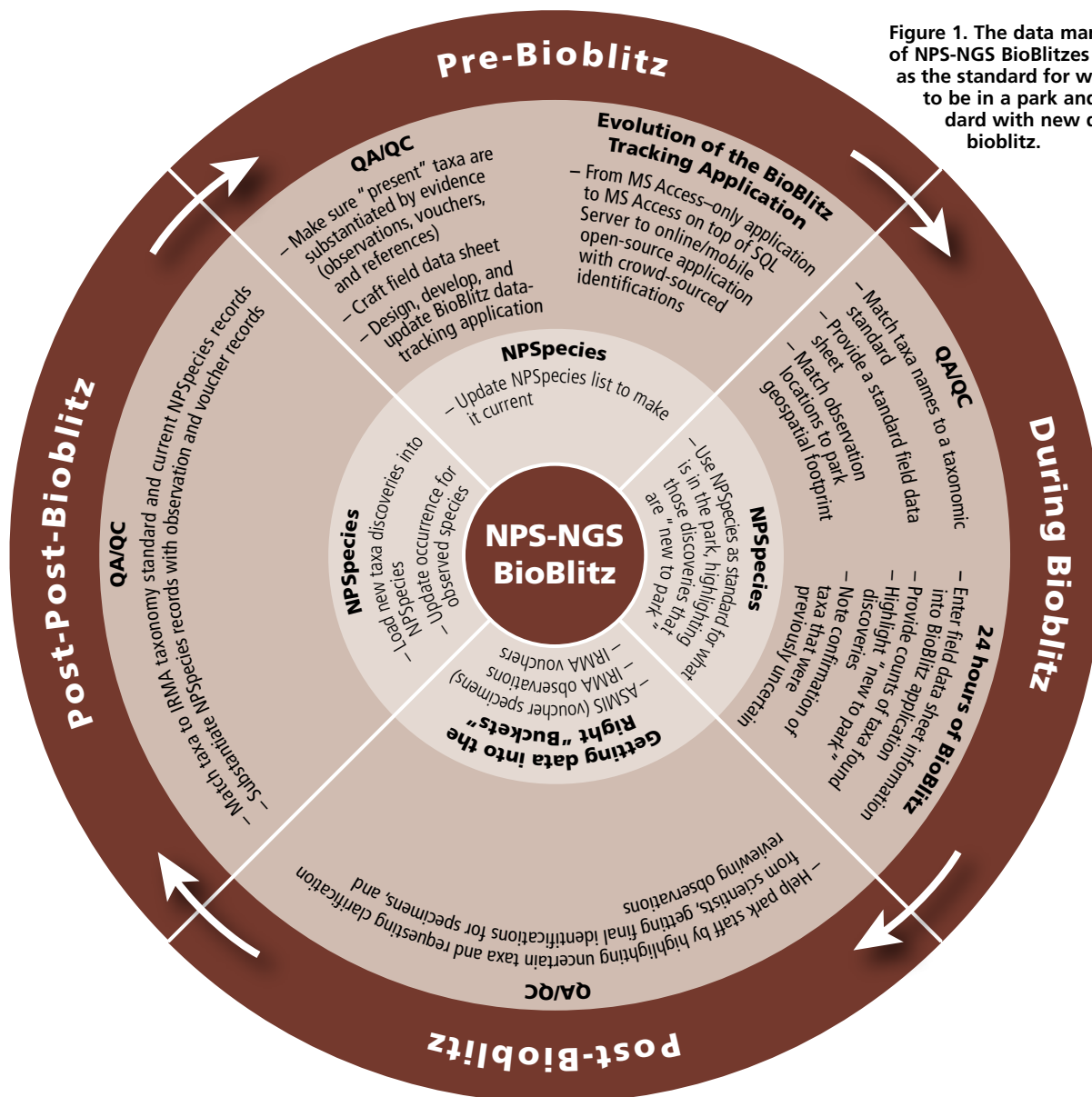
> Bioblitz events provide an opportunity for more than the enhancement of a park's NPSpecies inventory. The buildup to an NPS-NGS BioBlitz requires coordination by invited scientists who have been identified as being able to increase the understanding of often lesser-known taxonomic groups. Additionally, more than 3,000 people, including more than 1,500 schoolchildren, typically participate in a bioblitz and the concurrent biodiversity festival.

and research activities must be reviewed and added. In addition, the taxonomic nomenclature often needs to be updated to reflect the latest science. Once this is complete, the final step is to update the records to indicate how certain the park is that an organism occurs in the park.

QA/QC
In order to prevent errors from making their way into species lists from bioblitz-collected data, we take several quality assurance steps. This includes crafting a standardized field data sheet for use by

Figure 1. The data management life cycle of NPS-NGS BioBlitzes relies on NPSpecies as the standard for what species are known to be in a park and updates that standard with new discoveries after the bioblitz.

The diagram is a circular wheel with the central hub labeled **NPS-NGS BioBlitz**. The four outer quadrants are labeled:

Pre-Bioblitz
- **QA/QC**
  - Make sure "present" taxa are substantiated by evidence (observations, vouchers, and references)
  - Craft field data sheet
- **Evolution of the BioBlitz Tracking Application**
  - From MS Access–only application to MS Access on top of SQL Server to online/mobile open-source application with crowd-sourced identifications
  - Design, develop, and update BioBlitz data-tracking application
- **NPSpecies**
  - Update NPSpecies list to make it current

During Bioblitz
- **QA/QC**
  - Match taxa names to a taxonomic standard
  - Provide a standard field data sheet
  - Match observation locations to park geospatial footprint
- **24 hours of BioBlitz**
  - Enter field data sheet information into BioBlitz application
  - Provide counts of taxa
  - Highlight "new to park" discoveries
  - Note confirmation of taxa found previously uncertain
- **NPSpecies**
  - Use NPSpecies as standard for what is in the park, highlighting those discoveries that are "new to park"

Post-Bioblitz
- **QA/QC**
  - Help park staff by highlighting uncertain taxa and requesting clarification from scientists, getting final identifications for specimens, and reviewing observations
- **Getting data into the Right "Buckets"**
  - ASMIS (voucher specimens)
  - IRMA observations
  - IRMA vouchers
- **NPSpecies**
  - Load new taxa discoveries into NPSpecies
  - Update occurrence for observed species

Post-Post-Bioblitz
- **QA/QC**
  - Match taxa to IRMA taxonomy standard and current NPSpecies records with observation and voucher records
  - Substantiate NPSpecies records with observation and voucher records

---

all registered scientists, standardizing the taxonomy for all species lists, and requiring that species marked as "present" in a park are substantiated by evidence (e.g., voucher specimens or detailed observations). At this stage we also decide upon the observation-tracking method to be used, and further design and develop it as necessary. Once we have developed a field data sheet and decided on an observation-tracking method, we draft standard operating procedures, including detailed instructions, for their use.

Evolution of the bioblitz observation-tracking tools

Getting started (2007)
Bioblitz participants used paper data sheets to note observations made in the field, which were then entered into an electronic spreadsheet. After transferring the data, the paper data sheets were discarded, making it impossible to check entries. The NPSpecies list was not used as a source for correct spelling of species names.

Getting relational (2008–2009)
Though the paper data sheets continued to be used in the field, we recognized the importance of the original field records and began to retain them following bioblitzes. Also, a participating network data manager developed a relational desktop database that replaced the spreadsheets and removed the need to reenter survey team information for every

2014 Golden Gate National Recreation Area BioBlitz Field Datasheet

Turn in completed datasheets to your Scientist Check-In Location, or: Crissy Center, 1199 East Beach, Presidio San Francisco, CA 94129. If found, please call 415-426-5110

Inventory Leader Information
(1) Name　　　　　　　　　　　　　　(2) Primary Taxa:

(3) iNaturalist Username:　　　　　(4) Phone (during BioBlitz):　　　(5) Email:

(6a) Records added to iNaturalist in the field? All / Some / None　(6b) Account Used for Data Entry (Username):

(7) Date of Observations: ☐ March 28, 2014　☐ March 29, 2014　(8) Time Searching: From __ : __ AM / PM To __ : __ AM / PM

(9) Inventory Location: *BAKER CRISSY ELPO FOFU GIAC LANDS LOBOS MAHE Marine MORI MUBE MUWO PRSF-NW RANCHO*

Other: _____

(10) Location Description/Habitat Type:

(11) Sampling Method:

(12) Species Information	Qty. Obs.	Coordinates		Evidence	Comments
	actual # or est. # (≤25, ≤100, >100)	☐ Geographic, WGS84 (decimal degrees) † ☐ UTM Zone 10, NAD83 (meters) † decimal degrees preferred		enter all that apply: Observed Photo Specimen Audio rec. / if P, S, or A enter specimen ID / photo no. / audio file name	(include alternate names of species)
Scientific (preferred) or Common Name		X (easting, longitude)	Y (northing, latitude)		
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					
10.					
11.					

Figure 2. Standardized field data sheets ensure consistency of information collected during a bioblitz, capturing the who (inventory leader), what (species information), where (inventory location and coordinates), when (date of observations), and how (sampling method) of biodiversity discovery.

observation. Database forms allowed users to navigate through a menu system to look up official taxonomic names associated with the NPSpecies list for each park. By controlling these "pick lists" we avoided spelling errors and more easily saw when new species were added. Based on these improvements, we were able to provide counts of the number of different species found by taxonomic category (e.g., mammals, birds, vascular plants) during the bioblitz.

Tighter coupling with NPSpecies (2010–2011)
Field data sheets evolved to include a unique identification number displayed on each sheet, which made information easier to trace back to a team leader and easier to find in the database. Data managers scanned the field data sheets in order to make it easy for data entry technicians to review the original entries and to create digital records. The database application was split into two parts:

a simplified data entry tool that closely resembled the field data sheet and a more complex reporting tool. The reporting tool was based on a desktop NPSpecies application developed in 2005, allowing for even closer tie-in to the updated NPSpecies list for the park. New reports could be created, including a "Tree of Life" report, which interpretive staff have used to highlight extant and extinct species. Overall, these procedures facilitated tighter coupling with NPSpecies.

Going enterprise and getting social (2012–2013)
We continued to use paper field data sheets. The desktop database, however, was migrated to a higher-powered database system, which provided greater capacity for multiuser editing and better protection against accidental data loss. We designed a custom, menu-driven user interface that allowed bioblitz volunteers to enter the field data into the database. Naturalists then reviewed, matched, and

corrected scientific and common names and an administrator ran count reports. For the first time, Internet-based observation tools came into use, as Project Noah (www.projectnoah.org) observations from citizen scientists were imported into the bioblitz database and counted.

Going mobile (2014)
Paper field data sheets continued to be used in 2014 at the Golden Gate National Recreation Area bioblitz. In addition, we adopted the iNaturalist mobile application, which allowed citizens and scientists alike to contribute observations using their mobile devices. Rather than NPS

> Out of the 972 species identified at Biscayne National Park's bioblitz in 2010, 473 (49%) of the park total were found to be new to the official park species list.

staff designing and developing a database application for recording observations at the bioblitz, this time we selected an existing Web-based citizen science platform, iNaturalist (www.iNaturalist.org), to be the data repository. Staff worked to develop requirements for modifications to iNaturalist in order to make it compatible with the needs of the bioblitz. Investments in this observation tool not only allowed for greater participation by individuals who were not able to attend the 24-hour biodiversity discovery event, but also created a persistent resource for the park and citizens to continue to contribute species observations after the bioblitz.

The rush (during bioblitz)

Bioblitzes typically run from Friday morning through Saturday afternoon and usually start off slowly for data manage-

ment, as it takes a while for data sheets to arrive from the field. By the end of the day on Friday and especially on Saturday morning, things are hopping. Once groups start to return from the field, the data management team kicks into action, entering field data sheet information into

> Field guidebooks and access to authoritative Internet resources such as the Integrated Taxonomic Information System (ITIS) are critical in refining species observations during a bioblitz. These resources are used to correct misspellings and other common errors.

the bioblitz application, downloading and associating photos with the field records, and providing reports, such as counts of distinct species observed, to the bioblitz communication staff.

NPSpecies
By comparing species entered from the field data sheets with the computer-based NPSpecies list, we are able to highlight species that are new to a park and note when a previously uncertain species occurrence is confirmed. For example, out of the 972 species identified at Biscayne National Park's bioblitz in 2010, 473 (49%) of the park total were new to the official park species list.

QA/QC
NPSpecies serves as the source for scientific and common names when entering species from field data sheets, helping to prevent spelling errors. If a name is not found on the park's NPSpecies list, then a standardized taxonomy "lookup" or query of the database is used. If a name does not appear in the standardized taxonomy search results, then science volunteers try to validate the name with a recognized authority and follow up with field scientists. Field guidebooks and

access to authoritative Internet resources such as the Integrated Taxonomic Information System (ITIS) are critical in refining species observations during a bioblitz. These resources are used to correct misspellings and other common errors.

The unique identification number on each field data sheet is entered into the bioblitz database and saved into the scanned data sheet filename, making it easy to trace records from collection and entry to reporting. Data management volunteers also evaluate observation locations to be sure that they are within park boundaries.

The follow-up (post-bioblitz)

In the days and even years following a bioblitz, data management staff finalize data entry into NPS data systems, such as the Interior Collections Management System (ICMS), the Integrated Resource Management Applications (IRMA) Voucher system, and the IRMA Observations application.

NPSpecies
New species discoveries are added to the park's species list, updating the state of knowledge of what organisms occur in the park. Occurrence status is changed to "Present in Park" for species that were previously uncertain.

QA/QC
Data management staff work with park staff to finalize identification of specimens and review questionable identifications before adding these records to NPS data systems. They also work to ensure that NPSpecies records are substantiated with links to observations and vouchers.

Conclusion

The methods used to record species observations during National Park Service–National Geographic Society BioBlitzes have changed over time as new technologies have become available. The recent bioblitz at Golden Gate National Recreation Area in March of this year used an open-source, Web-based application that allowed contributions by citizens using applications on mobile devices. It also permitted crowd sourcing of identifications—that is, soliciting help from the online community of users.

As the technology used to document species observations at bioblitzes continues to evolve, NPSpecies remains the constant the National Park Service uses to communicate the depth of biodiversity in the parks. NPSpecies serves as the baseline for what species are known to occur in a park. It is used for quality assurance during a bioblitz to ensure that legitimate species names are associated with observations and to highlight discoveries of species that are new to a park. Finally, NPSpecies is updated after a bioblitz to reflect the new state of knowledge of what species are known to occur in a park.

About the authors

*Peter Budde is chief of the Restoration and Adaptation Branch, Biological Resource Management Division, Fort Collins, Colorado. **Simon Kingston** is a data manager with the NPS Inventory and Monitoring Division in Fort Collins, Colorado. They can be reached at peter_budde@nps.gov and simon_kingston@nps.gov.*